DOCUMENT RESUME

ED 442 834 TM 031 254

AUTHOR	Bridgeman, Brent; Burton, Nancy; Cline, Frederick
TITLE	Selecting Students with a General Reasoning Test (SAT I) or Tests in Academic Subjects (SAT II): Does It Matter for Non-Native Speakers of English?
PUB DATE	2000-04-00
NOTE	23p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 25-27, 2000).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*College Applicants; College Entrance Examinations; Higher Education; *Selection; Test Results
IDENTIFIERS	Nonnative Speakers; *Reasoning Tests; *Scholastic Assessment Tests

ABSTRACT

Using data from a sample of 10 colleges at which most students had taken both the SAT I: Reasoning Test and SAT II: Subject Tests researchers simulated the effects of making selection decisions using SAT II scores in place of SAT I scores. Students in each college were treated as forming the applicant pool for a more select college, and the top two-thirds of the students were selected using high school grade point average combined with either SAT I scores or the average of three SAT II scores. Included in the sample were 2,000 students whose best language was not English. These students had a better chance of being "admitted" with selections based on SAT II (subject test) scores than with selections based on SAT I (reasoning scores). Much of this advantage was the result of relatively high scores on language tests that could easily be part of the SAT II average. However, even when language tests were excluded from the SAT II average, non-native speakers were still slightly more likely to be selected with SAT II than SAT I. (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 442 834

Selecting Students with a General Reasoning Test (SAT I) or Tests in Academic Subjects (SAT II): Does It matter for Non-Native Speakers of English?

Brent Bridgeman
Nancy Burton
and
Frederick Cline

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

B. Bridgeman

1

Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, April, 2000

TM031254

BEST COPY AVAILABLE

Abstract

Using data from a sample of ten colleges at which most students had taken both the SAT I: Reasoning Test and SAT II: Subject Tests we simulated the effects of making selection decisions using SAT II scores in place of SAT I scores. Specifically, we treated the students in each college as forming the applicant pool for a more select college, and then selected the top 2/3 of the students using high school grade point average (HSGPA) combined with either SAT I scores or the average of 3 SAT II scores. Included in the sample were 2000 students whose best language was not English. These students had a better chance of being "admitted" with selections based on SAT II (subject test) scores than with selections based on SAT I (reasoning) scores. Much of this advantage was the result of relatively high scores on language tests that could be part of the SAT II average. However, even when language tests were excluded from the SAT II average, non-native speakers were still slightly more likely to be selected with SAT II than with SAT I.

The SAT I: Reasoning Test measures abilities in quantitative and verbal reasoning that develop over a long period of time and are not closely tied to specific academic subjects. The SAT II: Subject Tests "measure your knowledge and skills in a particular subject and your ability to apply that knowledge"(College Board, 1999). In terms of their overall ability to predict freshman grades, the SAT I and SAT II tests are nearly identical (Ramist, Lewis, & McCamley-Jenkins, in press). Nevertheless, because of their closer link to school subjects, though not to a particular well-specified curriculum, SAT II tests are sometimes seen as fairer and inherently less vulnerable to complaints of test bias. If a group does poorly on the SAT I, there is a tendency to blame the test, while poor performance on SAT II: Chemistry, for example, is more likely to be blamed on the quality of the chemistry instruction in the school. The University of California system recently increased the weight placed on SAT II tests relative to the weight on SAT I, and some institutions have proposed simply substituting SAT II for SAT I in admissions decisions. Such a substitution could impact not only the quality of the class selected, but also its composition.

The current research focuses on the consequences for language minority groups (and other minority groups) of substituting SAT II tests for the SAT I in the selection of a freshman class. Because the database to be used includes freshman grades, we can model not only the composition of the selected class, but also its academic success, at least to the extent that success can be defined by grades.

Method

Sample

Colleges in the sample were selected from a database of 23 colleges that was assembled for an SAT I validity study which compared the predictive validity of the old SAT to the new SAT I (Bridgeman, McCamley-Jenkins, & Ervin, 2000). This database contains SAT I and SAT II scores and responses on the Student Descriptive Questionnaire (SDQ), including high school grade point average, ethnic identification, best language (English, English and another, or another), parental education, family income, and intended college major. In addition, the database contains the freshman grade point average (FGPA). Students in the database were freshmen in 1995, so scores were available for relatively recent versions of the SAT II tests including Writing and Math IIC (advanced math that requires calculator use).

From this database we selected only colleges in which at least 80% of the freshman class had taken SAT II: Writing plus at least one other SAT II: Subject Test. Thus, at the campuses studied, students who took SAT II tests were the rule and not the exception. Of the 10 colleges included in the final sample, four were campuses of a public university system in the West, one was a private university in the Southeast, one was a private university in the Midwest, one was a private college for women in the Middle States, one was a private university in New England, and two were small private colleges in New England.

Responses to the Student Descriptive Questionnaire that students fill out when they register for the SAT were used for ethnic group identification. Responses to the question on best language were used to define the English as a second language (ESL)

sample. About 85% of the sample indicated only "English," 12% indicated "English and another," and 3% indicated only "Another." Students responding "English and another" or "Another" were placed in the ESL category.

Analyses

Freshmen at each of the 10 colleges who had scores on SAT II: Writing and at least one other SAT II test were treated as if they formed an applicant pool for an even more selective institution. At each college, 2/3 of the "applicant pool" was "selected" based on various score composites. A second set of analyses "selected" the top 1/3. Because any realistic selection scenario would include the high school grade point average (H), we decided to include H in each composite even though this would have the effect of muting the differences between selections made by alternative models. In each model, the composite score was formed by equally weighting each test score and giving H equal weight with the combined test scores. Prior research, with weights averaged over 685 within-college validity studies (with HVM), suggests that such weighting is close to optimal for predicting the freshman grade point average (Willingham, 1990).

The following composites were used:

(Note.—H=high school GPA; V=SAT I-Verbal; M=SAT I-Math; W=SAT II: Writing)

H+V+M (HVM)

H+Subject Test Average (H[SA])

H+Subject Test Average excluding language tests (H[SA-NL])

H+V+M+W (HVMW)

H+V+M+best Subject Test (HVMB)

H+V+M+best non-language Subject Test (HVM[B-NL])

Within each college, students selected by one of the new composites with SAT II scores were compared to students selected by the traditional HVM index. Freshman grade point averages (FGPA) were compared for four groups: 1) students selected by both the new composite and traditional index, 2) students rejected by both methods, 3) students selected by the new but rejected by the traditional, and 4) students selected by the traditional but rejected by the new. For these analyses, standardized differences in FGPA were computed within each college comparing the FGPA of students in Group 1 with students in Group 2 and comparing students in Group 3 with students in Group 4. These standardized differences (*d*) were weighted by the number of students in the relevant groups in each college and averaged across colleges.

Results and Discussion

Table 1 compares the FGPAs of students selected by HVM and those selected by H plus the average of the subject tests (H[SA]). Because of the high correlation between SAT I and the average of the SAT II tests (r = .84), and because H was used in both selection methods, the selection decision was the same under both models for 86% of the students. The comparison of Group 1 (students selected by both procedures) and Group 2 (students rejected by both procedures) suggests that valid selections can be made even though the initial selection pools were already quite restricted because they consisted of only students who had already been admitted to and enrolled in selective colleges. Grades of the students "selected" by both methods were about 0.8 SD units higher than grades of students not selected (or about 0.4 points higher on the typical 0-4 GPA scale). The comparison of Groups 3 and 4 (the 14% of the students who were selected by one method but rejected by the other) indicated that freshman grades of students in these two

groups were nearly identical. Thus, with freshman grades as the criterion, there is no reason to favor either SAT I or SAT II in making selection decisions. The same was true for the comparison that excluded language tests from the Subject Test average.

As expected, there was even more overlap in the models that added subject tests to V and M rather than replacing V and M. In the model that added Writing, 93% of the selection decisions were the same as with HVM alone. Similarly, for HVMB, 93% were identical, and for HVM(B-NL), 94% were identical. There were no significant differences in any of the comparisons between the FGPAs of the groups admitted by one model and rejected by the other.

Although the overall success of students selected using SAT I is comparable to the success of students selected using SAT II, there might still be differences in the ethnic or gender composition of groups selected by the different criteria. Figure 1 shows the percent of the pool of female students that was selected by being in the top 2/3 for each selection index. Each percentage was slightly below the 66.7% that would be expected if there were no gender differences on any of the selection instruments. Although there was relatively little variation among the various indices, including SAT II: Writing along with SAT I scores increased the percentage of women selected by a small but statistically significant 2.5 percentage points (standard error of each percentage is about 0.7).

As indicated in Figure 2, more substantial differences were evident in the ethnic group comparison of HVM selections with selections that combined H with the average of the Subject tests (H[SA]) and selections that combined H with the average of the non-language subject tests (H[SA-NL]). In particular, the proportion of Mexican American and Other Latino students selected would increase if H(SA) were used in place of HVM.

BEST COPY AVAILABLE

Because we were keeping the size of the admitted class fixed, at least one of the other groups had to show a reduction in this zero-sum game. Numerically, the loss of White and Asian American students balanced the gains of Mexican American and Latino students, although the percentage loss in each of these groups was small because of the relatively large numbers of White and Asian American students in the sample. The percentage of the eligible African American group that was selected was virtually identical with either model.

When language tests were excluded from the subject test average, the advantage for the Mexican American and Other Latino groups essentially disappeared; the small apparent advantage remaining was not statistically significant (standard errors of 1.6 and 2.2 respectively for the percentages in the Mexican American and Other Latino groups). The impact of including or excluding the language tests is somewhat muted because only 43% of the Mexican American students and 51% of the Other Latino students took one of the Spanish subject tests (either Spanish or Spanish with Listening). In order to gauge the impact of the language test on the likelihood of selection, we examined the sample of Mexican American and Other Latino students who had taken one of the Spanish tests. As shown in Figure 3, in both groups almost twice as many students were selected with the index including the subject test average as by the index that used V and M scores. Excluding the Spanish test from the Subject test average markedly reduced the number of students selected from these groups. Recall that half of the weight in the prediction equation is on the high school average and the Spanish test is approximately one-third of the subject test weight (or 1/6 of the total weight); given this relatively small weight, the effect of including or excluding the Spanish test is indeed dramatic. Test means show the

reasons for this relative advantage. In the combined Hispanic groups, the mean score on the Spanish test was 147 points higher than the mean score on SAT I: Verbal (666 vs. 519 ; SDs 90 and 91 respectively); in the White sample, the mean score on the Spanish test was 85 points lower than the mean score on SAT I: Verbal (556 vs. 641; SDs 89 and 77 respectively).

The selection index that included the average of the subject tests resulted in the selection of more Hispanic students than selections based on HVM. We next determined how successful these students were, defining success as achieving a freshman GPA of 2.5 or better, and again using the sample of students who had taken one of the Spanish subject tests. For the sample of Mexican American students, including those who were not selected with any of the indices, 59% were successful by this criterion. (We also investigated a GPA of 2.0 or better as the criterion, but the overall success rate was 87%, allowing for little variation among the different selection methods.) In the Other Latino sample, the overall success rate was 69% for the 2.5 or better criterion. As indicated in Figure 4, the students selected by HVM were most successful on a percentage basis; 79% of the Mexican American students selected by HVM were successful compared to 66% for H(SA). For the Other Latino students, 84% selected by HVM were successful compared to 76% for H(SA). If maximizing the percent of successful students in the Hispanic groups were the goal, selections should be based on HVM. However, recall that many more Hispanic students were selected with H(SA) than with HVM. If emphasis is placed on the number of successful students selected from the subgroup instead of on the percent of students in the selected subgroup who are successful, a different conclusion is reached. As indicated in Figure 5, the number of successful Hispanic students was

greatest for selections based on the index that used the average of the subject tests, including the Spanish subject test. If admitting the maximum number of potentially successful Hispanic students were the goal, selections should be based on H(SA).

Selecting the Top 1/3

The above analyses assumed that, within each institution, 2/3 of the class would be selected. The following analyses were based on selecting the top 1/3 within each institution. For the three primary selection models (HVM, H[SA], and H[SA-NL]), the proportion of women selected was the same, 31%. Because the pool contained slightly more women than men (7610 to 6246), the number of women selected was almost the same as the number of men selected (for H[SA], 2376 women and 2284 men were selected). As with the top 2/3 selection, adding SAT II : Writing to HVM yielded an increase of about 2 percentage points to the percentage of women selected (from 30.7% to 32.6%), though 93% of the selections are the same with HVM as with HVMW. Figure 6 shows the percent selected from each ethnic group for each of the three major indices. The general pattern is the same as was observed for the top 2/3 selections; for all of the selection models, White students were overrepresented, Asian students were proportionally represented, and the other groups were underrepresented relative to their numbers in the applicant population. Mexican American, Other Latino, and ESL students were somewhat more likely to be admitted with the model that used the average of the subject tests than with the model that used SAT I: V and M scores.

Success rates, defining success as achieving a grade point average of at least 2.5, were comparable across the different selection models. Table 2 shows the number of students selected by both methods, rejected by both methods, and selected by one but

BEST COPY AVAILABLE

rejected by the other. For each of these groups, the percent of the selected students who were successful is also shown. Success rates for students in Group 3 (selected by H[SA] and not by HVM) were virtually identical to success rates in Group 4 (selected by HVM but not H[SA]), except in the two Hispanic groups in which success rates were slightly higher for the HVM selections. Although ESL students were much more likely to be selected with H(SA) than with HVM, success rates among those selected by either index were the same. The relatively low success percentages in Group 2 (rejected by both HVM and H[SA]) is evidence for the validity of selections based on high school average and either SAT I or SAT II test scores.

Conclusion

Colleges that are selecting students from applicant pools that are similar to the enrolled students in this study could select an equally qualified class using the SAT II: Subject Tests as they could using the SAT I: Reasoning Test. Switching to the SAT II test average would have a minimal impact on the number of women or African American students selected. ESL students generally would have a higher likelihood of being selected with SAT II. Similarly, noticeably more Mexican American and Other Latino students would be selected with the subject test average, especially if students could submit the Spanish or Spanish with Listening subject tests. Adding the SAT II: Writing test to the SAT I: Reasoning Test would increase the proportion of women selected, but by less than 3 percentage points.

All of the institutions in the current sample were at least moderately selective, and most were highly selective. Because the pool we used consisted of only students who had already been through a rigorous selection process, even our "rejected" students were

a relatively elite group. Further study is needed before generalizations to less selective institutions and more diverse applicant pools could be made.

References

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). Predictions of freshman grade point average from the revised and recentered SAT I: Reasoning Test. (College Board Report No. 2000-1). New York: College Entrance Examination Board.

College Board (1999). Taking the SAT II: Subject Tests. New York: College Entrance Examination Board.

Ramist, L. Lewis, C., & McCamley-Jenkins, L. (in press). Using Achievement Tests/SAT II: Subject Tests to demonstrate achievement and predict college grades: Sex, language, ethnic, and parental education groups. Princeton, NJ: Educational Testing Service.

Willingham, W. W. (1990). Introduction: Interpreting predictive validity. In W. W. Willingham, C. Lewis, R. Morgan, & L. Ramist, Predicting college grades: An analysis of instututional trends over two decades (pp. 3-21). Princeton, NJ: Educational Testing Service.

Table 1

FGPAs for Students Selected or Rejected in Top 2/3 by HVM and/or H(SA)

College	Group 1 (in both)			Group 2 (out both)			*d (1-2)*	Group 3 (in H[SA] only)			Group 4 (in HVM only)			*d (3-4)*
	n	M	SD	n	M	SD		n	M	SD	n	M	SD	
1	686	3.33	0.43	291	3.05	0.47	**0.62**	87	3.25	0.43	86	3.17	0.48	**0.18**
2	658	3.04	0.66	293	2.54	0.73	**0.72**	60	2.69	0.63	63	2.63	0.73	**0.09**
3	191	3.34	0.46	78	3.04	0.43	**0.67**	30	3.27	0.31	30	3.19	0.41	**0.22**
4	162	3.27	0.48	76	2.74	0.51	**1.07**	15	3.13	0.33	15	2.97	0.35	**0.47**
5	181	3.16	0.59	80	2.52	0.81	**0.90**	28	2.96	0.32	28	2.96	0.29	**0.00**
6	586	3.57	0.44	268	3.32	0.46	**0.89**	73	3.57	0.39	72	3.47	0.32	**0.29**
7	1406	3.01	0.59	632	2.51	0.62	**0.83**	153	2.72	0.63	155	2.69	0.54	**0.05**
8	1421	3.09	0.51	649	2.72	0.51	**0.73**	198	2.86	0.48	189	2.88	0.49	**-0.04**
9	1676	3.21	0.49	764	2.65	0.55	**1.08**	168	2.80	0.56	159	2.90	0.51	**-0.19**
10	1267	2.91	0.62	550	2.50	0.60	**0.67**	177	2.60	0.59	171	2.59	0.66	**0.02**
						Wtd. Avg.	**0.82**						Wtd. Avg.	**0.03**
						Std. Error of Wtd. Avg.	**0.06**						Std. Error of Wtd. Avg.	**0.05**

Table 2

Success (GPA 2.5 or greater) Rates for Students Selected or Rejected in top 1/3 by HVM and/or H(SA)

Group	Group 1 (in both)		Group 2 (out both)		Group 3 (in H[SA] only)		Group 4 (in HVM only)	
	n	% GPA 2.5+	n	% GPA 2.5+	n	% GPA 2.5+	n	% GPA 2.5+
Men	1869	90	3495	69	415	85	485	84
Women	1860	95	4755	77	516	88	479	88
African Am.	34	94	437	70	16	88	13	85
Asian Am.	1239	91	2816	69	317	87	353	81
Mex. Am.	86	83	781	61	30	73	26	88
Other Latino	77	96	415	67	32	78	25	88
White	1983	93	3246	81	377	89	480	88
ESL	349	90	1362	66	200	80	89	80

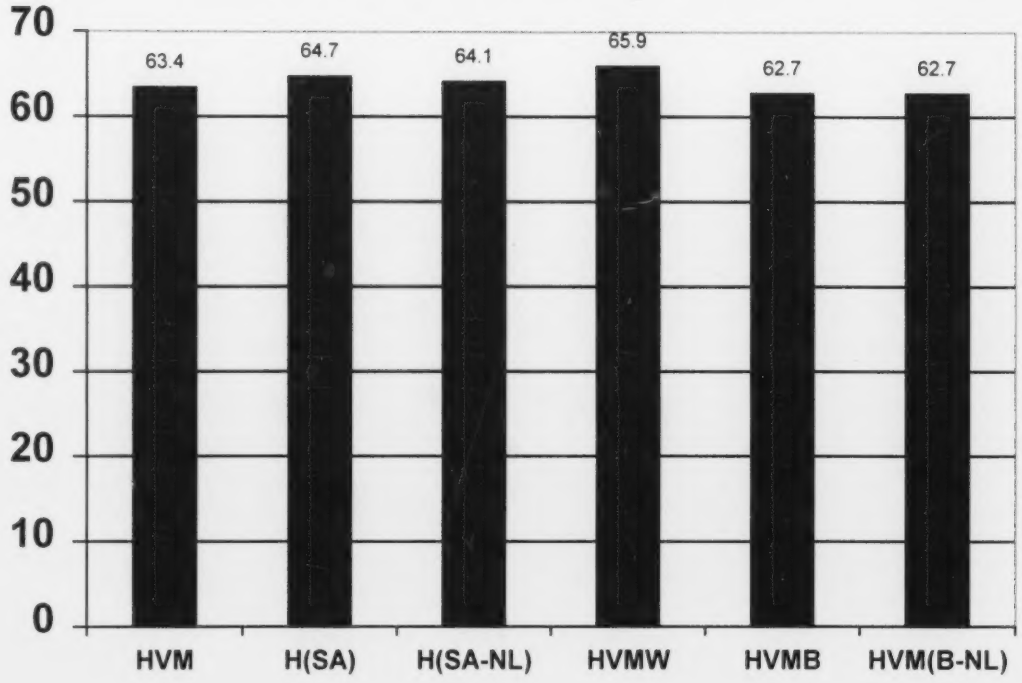

Figure 1. Percent of Females in Top 2/3 Selected Under Six Alternative Models

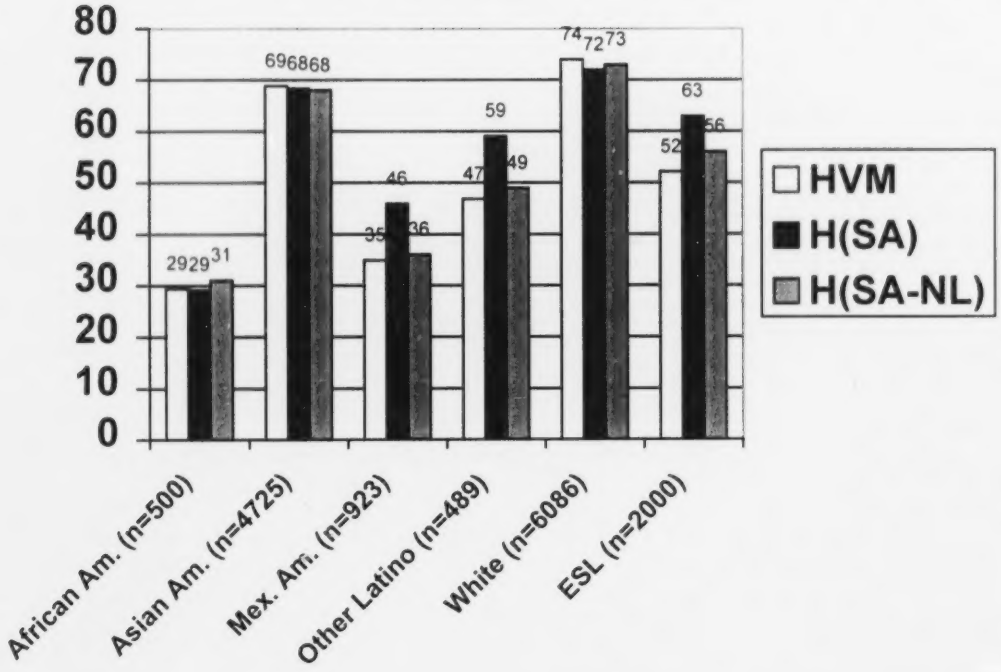

Figure 2. Percent of Each Group Selected by HVM, H(SA), and H(SA-NL) for upper 2/3

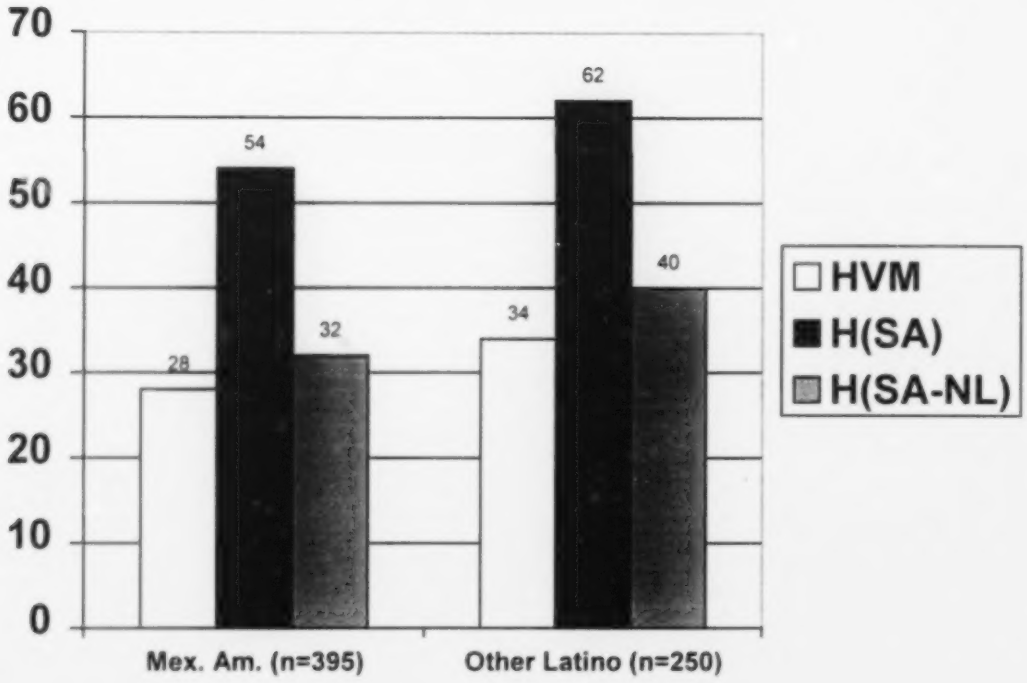

Figure 3. For Students Who Took a Spanish Subject Test, Percent of Each Group Selected by HVM, H(SA), and H(SA-NL) for upper 2/3

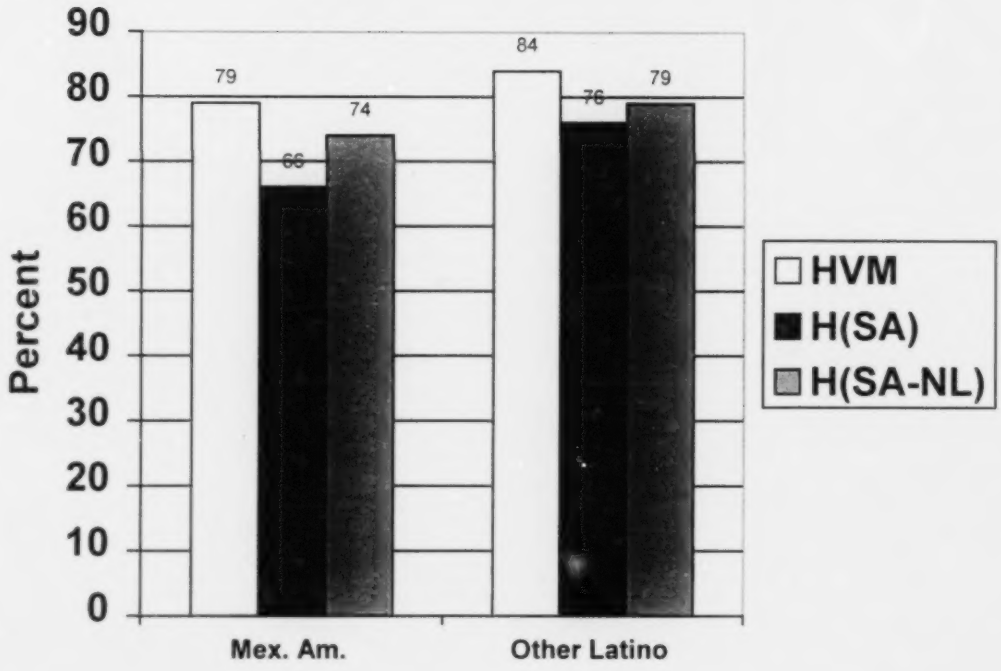

Figure 4. For Students Who Took a Spanish Subject Test, Percent of Selected Students in Each Group Who Are Successful (GPA 2.5 or higher).

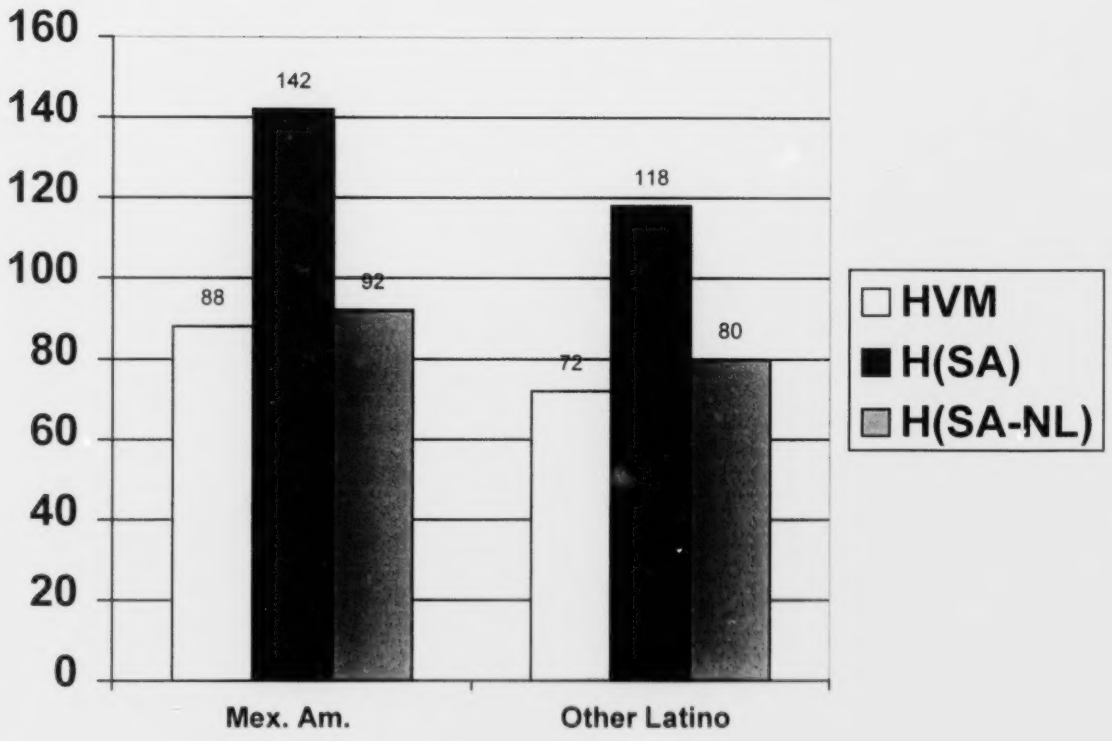

Figure 5. For Students Who Took a Spanish Subject Test, Number of Selected Students in Each Group Who Are Successful (GPA 2.5 or higher).

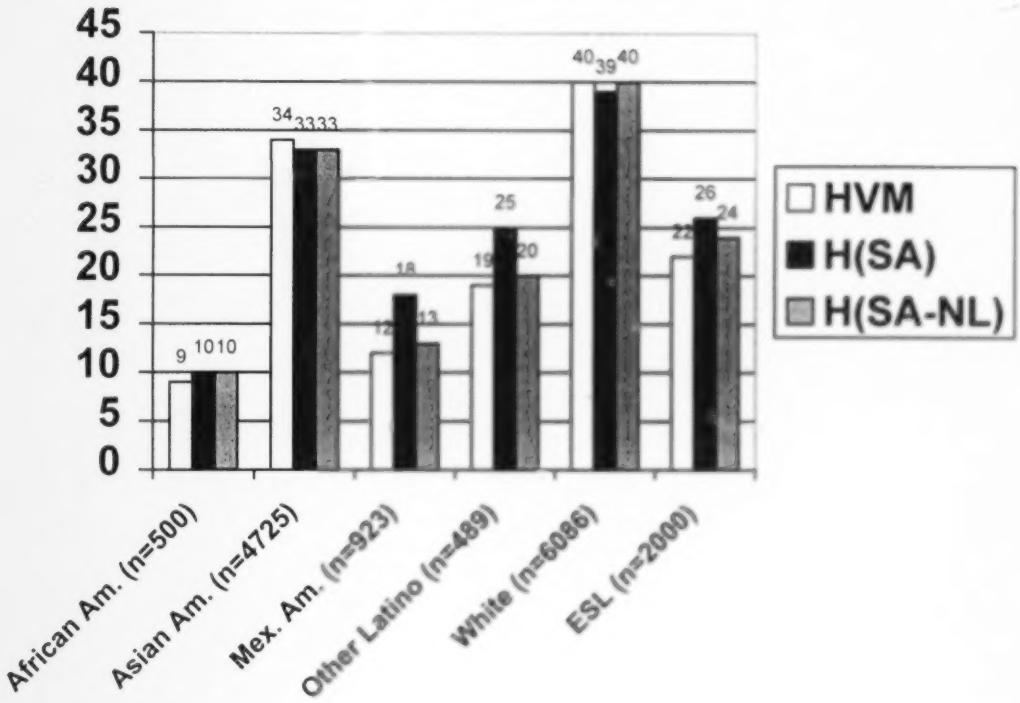

Figure 6. Percent of Each Group Selected by HVM, H(SA), and H(SA-NL) for upper 1/3.